Tautomerization constrains the accuracy of codon-anticodon decoding

Andriy Kazantsev* and Zoya Ignatova*

Institute of Biochemistry and Molecular Biology, Department of Chemistry, University of Hamburg, Hamburg, Germany

E-mail: kazantsev.and@gmail.com; zoya.ignatova@chemie.uni-hamburg.de

Abstract

G∘U(T) mismatch has the highest contribution to the error rate of base pair recognition in replication, as well as in codon-anticodon decoding in translation. Recently, this effect was unambiguously linked to keto-enol tautomerization, which enables the Watson-Crick (WC) geometry of the base pair. Structural studies of the ribosome revealing G∘U in the WC geometry in the closed state of the A-site challenge the canonical induced-fit model of decoding and currently lack a physicochemical explanation.

Using computational and theoretical methods, we address effects of the ribosomal A-site on the wobble↔WC tautomerization reaction in G∘U (wb-WC reaction), and the consequent implications for the decoding mechanism in translation. The free energy change of the wb-WC reaction in the middle codon-anticodon position was calculated with quantum-mechanical/molecular-mechanical umbrella sampling simulations. The wb-WC reaction was endoergic in the open A-site, but exoergic in the closed state. This effect can be explained in part by the decreased polarity of the closed A-site.

We developed a model of initial selection in translation that incorporates the wb-WC reaction parameters in the open and closed states of the A-site. In the new model,

the exoergic wb-WC reaction is kinetically restricted by the decoding rates, which explains the observations of the WC geometry at equilibrium conditions. Moreover, the model reveals constraints imposed by the exoergic wb-WC reaction on the decoding accuracy: its equilibration counteracts the favorable contribution from equilibration of the open-closed transition. The similarity of the base-pair recognition mechanism in DNA polymerases allows extending this model to replication as well. Our model can be a step towards a general recognition model for flexible substrates.

INTRODUCTION

The accuracy of base pair recognition during replication and translation has a fundamental role in cell fitness and evolution[1–4]. Some mismatches, such as G∘U (G∘T) mismatch, have particularly large contributions to error rate[5–10]. The tautomeric hypothesis by Watson and Crick[11],[12] suggests that a source of base pair recognition errors can be attributed to keto/enol and amino/imino tautomerism of nucleobases. G∘U(T) base pair formed with the canonical keto states of the nucleobases is unable to adopt the Watson-Crick (WC) geometry and predominantly exits in the wobble (wb) geometry[13]. Formation of enol tautomeric states (G* or U*(T*)) enables the WC geometry, potentially leading to errors during base pair recognition, but the biologically-relevant mechanisms of this process remained elusive until recently.

Using computational chemistry methods, Brovarets' and Hovorun predicted wobble↔WC tautomerization reaction in G∘U(T) base pair[14,15] (wb-WC reaction, Fig. 1A), which was experimentally confirmed by NMR in DNA and RNA duplexes in water solution[13], where this reaction was endoergic (i.e. the wobble geometry was more populated than the WC)[13,16]. Recently, by incorporating tautomeric populations measured in DNA duplexes in water solution into a numerical kinetic model of DNA replication, a good agreement between predicted and experimentally observed misincorporation rates was obtained[16]. It might indicate a negligible role of the active site environment on the properties of the wb-WC reaction. However,

both experimental[17] and computational[18,19] studies demonstrate stabilization of the WC geometry of G∘T in the closed state of active site in several DNA polymerases, implicating a more complicated mechanism of base pair recognition. The WC geometry of G∘U was also observed in the closed state of the decoding site (A-site) of the ribosome at the first two codon-anticodon positions: in crystal structures[20–23] and later in Cryo-EM structures[24].

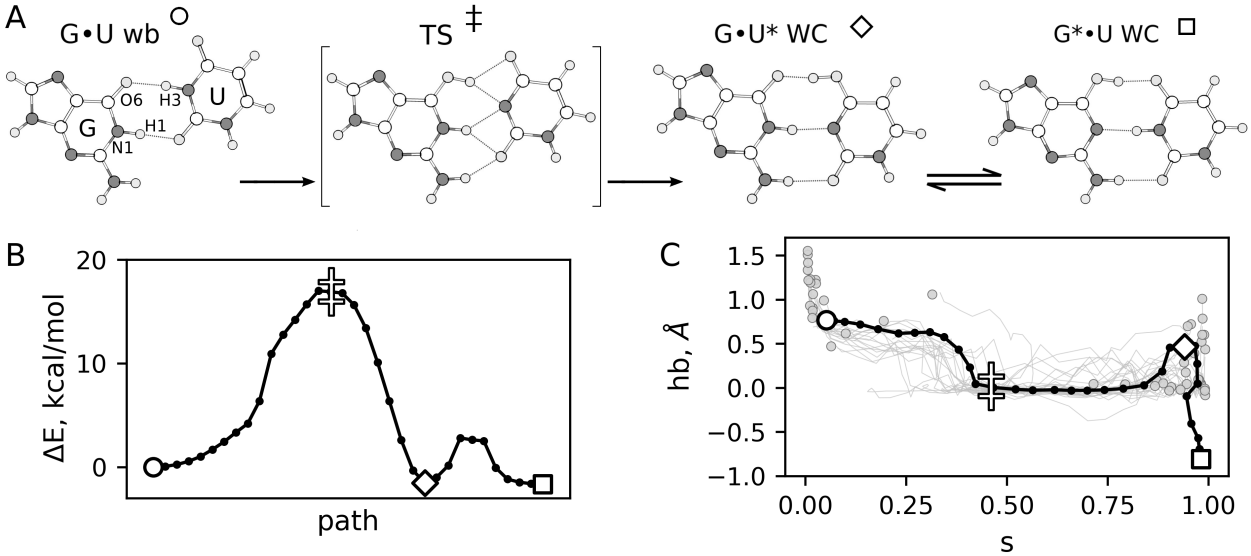


Figure 1: The wb-WC reaction. **A** – local minima and the transition state (TS) along the wb-WC reaction. Asterisk denotes enol tautomers. The markers shown at each state are used to denote the corresponding states in **B** – **C**. The wb-WC reaction consists of slow tautomerization reaction from G∘U wb to G∘U* WC, followed by fast double proton transfer reaction between G∘U* and G*∘U. **B** – reference potential energy profile of the wb-WC reaction calculated with nudge elastic band method (NEB) in gas phase. **C** – reference wb-WC reaction path from NEB, projected on the collective variables used in umbrella sampling calculations s (path collective variable) and hb (proton transfer variable). Gray lines and gray circles denote trajectories from the committor analysis, initiated from the TS

These observations of the WC geometry of G∘U(T) challenge the classical model of substrate recognition in replication and translation – the induced-fit model, which assumes rigid substrates. The open↔closed conformational transition of the ribosomal decoding site, the molecular mechanism of the "induced fit"[24,25], provides the highest decoding accuracy in the equilibrium limit, but is shifted from equilibrium due to its high forward rates, and the rates of the following irreversible reaction[26–29]. Theoretical studies based on the rigid substrate approximation have concluded that decoding in translation has been evolutionarily

optimized for higher rate rather than for higher accuracy[27,28,30]. Since the rate constants affecting the open↔closed transition are not rate-limiting in translation elongation[27], a possible advantage of such trade-off remains unclear and additionally challenges the induced-fit model. Hovorun and coworkers have suggested a non-equilibrium exoergic wb-WC reaction in the closed state of a DNA polymerase active site, kinetically restricted by the lifetime of the recognition state[15]. Such mechanism applied to the codon-anticodon decoding might solve the above described problems, but requires formalization and detailed analysis.

Clearly, studying environmental effects of the ribosomal decoding site on the wb-WC reaction is essential for understanding the decoding mechanism. In the absence of experimental methods suitable for addressing this problem, computational methods provide such opportunity. Until now, two computational studies have addressed nucleobase tautomerism in codon-anticodon decoding[31,32], both of which argue against stabilization of the WC geometry in G∘U at the decoding site, thus contradicting the experiments[20–24]. However, these computational studies used force fields for energy calculations, which may not be the most accurate approach to study tautomerization, especially when the rare tautomers were not parameterized, as in ref.[31].

Here, we overcome the force-field-related biases by performing hybrid quantum-mechanical/ molecular-mechanical (QM/MM) simulations. We use QM/MM umbrella sampling (US) method to calculate free energy profiles of the wb-WC reaction in benchmark systems (DNA duplex in water, and a single base pair in benzene), ribosomal A-site models and active sites of DNA polymerases. Our calculations reveal positive free energy of the WC geometry formation ΔG_{wc} in the model of the open A-site, and negative ΔG_{wc} in the closed model, consistent with the structural studies[20–24]. We also found different effects of DNA polymerases on the wb-WC reaction: ΔG_{wc} in the active site of polymerase β was negative, consistent with the structural studies[17], while ΔG_{wc} in DNA polymerase T7 was not significantly affected compared to the DNA duplex in water. We analyze dielectric constant of the studied environments to explain the observed effects on the wb-WC reaction.

To study implications of the exoergic wb-WC reaction in the closed A-site on the codon-anticodon decoding, we incorporated the wb-WC reaction into the model of initial selection in translation and derived analytical solutions for the new model. In this model, the exoergic wb-WC reaction is kinetically restricted by the residence time of the closed state of the decoding site. The new model is consistent with previously reported WC geometry of G∘U in the structural studies at equilibrium conditions[20–24]. In addition to uniting structural and kinetic data, the model reveals constraints on decoding: the equilibration of the wb-WC reaction counteracts the equilibration of the open↔closed transition, thereby constraining the error rate of decoding.

The similarities with DNA polymerases suggest that the proposed model can be also applied to replication. The model derived here might be a step towards a more general model of recognition capable of describing flexible substrates.

METHODS

Details on the methods are provided in Supplementary Methods.

QM calculations. We performed quantum chemistry (QM) calculations to obtain the reference dependence of the wb-WC equilibrium on the dielectric constant of a solvent, and to select the optimal QM level of theory for our QM/MM calculations. Using geometries of the wb-WC reaction minima and transition state (TS) optimized at density functional theory (DFT) levels ωB97X-D3/def2-TZVP, B3LYP-D3BJ/def2-TZVP and BLYP-D3BJ/def2-SVP in gas phase and implicit water model, we calculated single-point energies with multiple DFT and semiempirical (SE) methods. The methods were assessed based on the energies (ΔE_{wc}) or enthalpies (ΔH_{wc}) of the WC geometry formation ($\Delta E_{wc} = \Delta E$(G*∘U WC) - ΔE(G∘U wb)). Our benchmark calculations showed PM7[33] as an optimal QM method for our QM/MM approach, as it was able to accurately reproduce ΔH_{wc} calculated with DFT (Table S2), while allowing vast sampling due to its low computational cost. The activation

enthalpies obtained with PM7 were much less accurate (Table S2). Orca 4.2.1[34] was used for all *ab initio* QM calculations, while MOPAC-2016[35] was used for all SE calculations.

System setup and MD simulations. We studied the environmental effects on the wb-WC reaction in the following systems: DNA heptamer in water solution and a single G∘T base pair in benzene (benchmark systems); ribosomal A-site models: closed, abasic and open; human DNA polymerase β (pol-β) and DNA polymerase from T7 virus (T7-pol).

DNA heptamer model was prepared from the solution NMR structure of DNA dodecamer (PDB ID: 1BJD), centered on one of the G∘T wobble base pairs in the dodecamer. Benzene model was prepared by solvating a single G∘T base pair from the DNA heptamer system in a benzene box.

pol-β model was prepared from the crystal structure of pol-β with a G∘C base pair in the closed active site (PDB ID: 4KLF)[36], in which we replaced dCTP with dTTP. T7-pol was prepared in a similar way using the crystal structure with PDB ID 1T7P[37] as the initial coordinates.

To prepare the A-site models, we used the crystal structure of *T.thermophilus* 70S ribosome bound to tRNA^{Thr} in the A-site (PDB ID: 6GSK) with U∘G base pair in the second codon (AUC) - anticodon (GGU) position[23]. All residues within 35 Å radius of the second codon nucleotide were selected for the model (Fig. 2A). The outer shell of the selected sphere was restrained in all simulations. In the "closed" (native) A-site model the rRNA residues A1492, A1493 and G540 were in *out* conformation, surrounding the codon-anticodon helix (Fig. 2B). In the "abasic" model the nucleobases of A1492, A1493 and G540 were deleted. In the "open" model, used only in QM/MM US simulations (see below), we used harmonic biases from Zeng et al.[32] to restrain A1492 and A1493 to the *in* position. All Mg^{2+} ions from the ribosome crystal structure were deleted in the models.

Unless specified otherwise, all models were solvated in TIP3P[38] water box with 0.15 M of NaCl. We employed a standard equilibration protocol (see Supplementary Methods), comprising steepest-descent minimization, solvent equilibration in NPT ensemble at standard

conditions and gradual heating of the solutes to 298 K. Production MD simulations were performed in NVT ensemble using Langevin thermostat with 2 fs integration step. All MD simulations were performed using periodic boundary conditions. CHARMM36 force field[39–41] in NAMD 2.12[42] was used for all classical MD simulations.

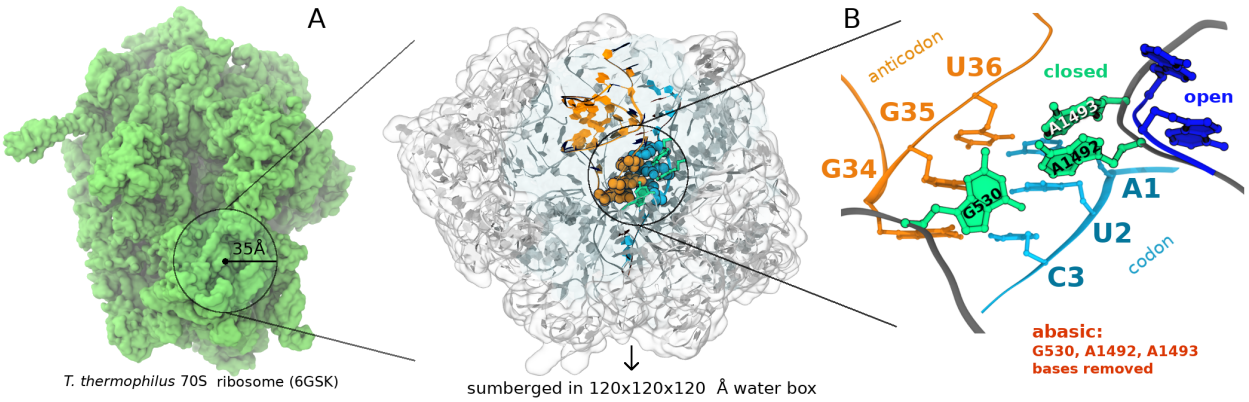


Figure 2: Preparation of the A-site models. **A** – visualization of the A-site model (right) and its approximate position and size in the 70S ribosome (left). **B** – close-up view on the codon-anticodon helix in the decoding center. The closed and the open A-site models differ by conformations of the rRNA residues A1492 and A1493 (green and blue in the closed and open states, respectively). In the abasic A-site model nucleobases of A1492, A1493 and G540 were deleted, leaving only the sugar-phosphate backbone.

Dielectric constant calculations. To analyze dielectric properties around the base pairs in the simulated models, we applied Kirkwood-Fröhlich formula (KFF), which relates dielectric constant (ε) to dipole moment fluctuations in a selected volume[43]. Using classical MD trajectories, we calculated ε of all nucleic acid, protein and water residues in spheres of radii ranging from 5 to 12 Å centered on the base pair of interest. Since ε calculated in a TIP3P water box using this approach depended on the probe volume radius until its convergence at $\sim$ 20 Å (Fig. S3), the approach only allows a qualitative comparison between the studied systems.

QM/MM US simulations. We applied QM/MM umbrella sampling simulations (US) to calculate the free energy profiles (potential of mean force, PMF) of the wb-WC reaction in the models of ribosomal A-site and active site of DNA polymerases, as well as benchmark systems. All QM/MM calculations were performed on PM7/CHARMM36 level. The QM

region comprised only nucleobases in a G∘U(T) base pair of interest, with the QM-MM boundary placed at the glycosidic bonds. To select collective variables (CV) for the US calculations, we first optimized the wb-WC reaction path in G∘U in gas phase using nudged elastic band (NEB) method at B3LYP-D3BJ/def2-TZVP level (Fig. 1B-C). The coordinate frames of this path became the reference for path collective variables[44] (pathCV), which were used to calculate the position on the path (s) of the wb-WC reaction in the US simulations. s variable described the geometry change from wobble to WC, while a distance difference $hb =$ d(O6-H3) - d(N1-H1) was used to describe proton transfers during the reaction. Selected CVs clearly distinguished all three local minima of the wb-WC reaction (Fig. 1C). Committor analysis was performed to verify the TS (Fig. 1C). Using s and hb as CVs, and adding a boundary potential to prevent large deviations from path z, we collected 100-400 ps of QM/MM US trajectories for each of 48 US windows per each system (∼120 ns in total) and calculated PMF using weighted histogram analysis method (WHAM). We focused only on the relative energies and positions of local minima in the wb-WC reaction in our US simulations

Kinetic modeling. Decoding rate constants for a codon-anticodon combination with a G∘U mismatch in the first position were obtained earlier by Pape et al.[45], but only in the low-fidelity buffer conditions, in which the accuracy of the initial selection was extremely low (1.1), contrasting the *in vivo* measurements[5–9]. Therefore, for our kinetic modeling we selected decoding rate constants measured at high-fidelity conditions, which result in more realistic accuracy[46,47]. In the absence of high-fidelity measurements corresponding to a G∘U mismatch, we used the set of rate constants corresponding to the A∘C mismatch in the first codon-anticodon position[46,47] Table S3. Rate constants k_2, q_3^c and q_3^{nc} were assigned arbitrary. Numerical solutions of the kinetic model were obtained by numerical integration of ordinary differential equations (ODE) in Python. We describe derivations of analytical solutions in Appendix.

RESULTS AND DISCUSSION

Dielectric constant of base pair environments

It was speculated previously that closing of the ribosomal A-site, via conformational change of the rRNA residues A1492, A1493 and G540, desolvates the codon-anticodon helix, increasing base-pairing selectivity by an energy penalty from the lost H-bonds with water in mismatches[25,48]. We hypothesized that desolvation may also bring an opposite contribution to the accuracy of G∘U mismatch recognition. Our benchmark calculations (Fig. S1) demonstrate that the wb-WC reaction is exoergic in gas phase and very nonpolar implicit solvents, corroborating previous computational studies[14,15,18,49]. We reasoned that a potentially decreased polarity of the closed ribosomal A-site may be responsible for the stabilization of the WC geometry of G∘U observed in the structural studies.

We applied Kirkwood-Fröhlich formula (KFF) to qualitatively compare dielectric constant ϵ of the decoding site environments between the closed and abasic models. In the latter model, nucleobases of residues A1492, A1493 and G540 were deleted (Fig. 2B). We observed a consistent ε decrease in the closed model for all codon-anticodon positions (Fig. 3). However, only at some distance cutoffs and positions the difference was statistically significant, which can be explained by the limited length and replica numbers of the MD trajectories (Fig. S4). From our KFF calculations we conclude that the A-site closing decreases ε of environment of all three codon-anticodon positions.

We also compared the active sites of pol-β and T7-pol using the same approach. The active site environment in pol-β was generally less polar compared to T7-pol, but the difference was statistically significant only at 12 Å cutoff (Fig. 3).

In sum, our KFF calculations revealed decreased polarity in the closed compared to the abasic ribosomal A-site model, and in pol-β compared to T7-pol.

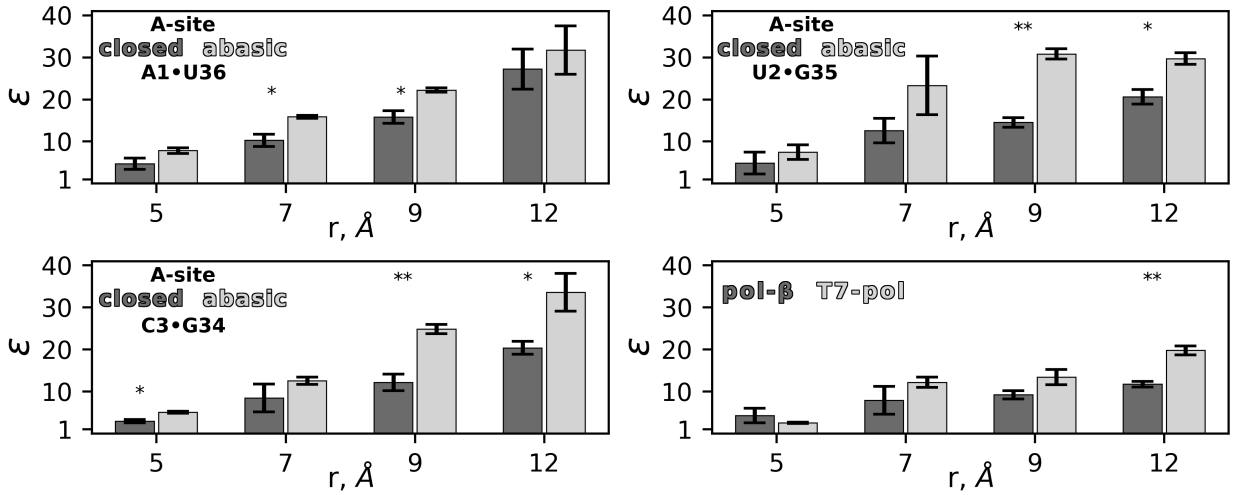


Figure 3: Dielectric constant ε of the base pair surroundings. **A**–**C** – ε difference between closed and abasic A-site models around each codon-anticodon base pair (A1-U36, U2-G35 and C3-G34) measured in spheres of radius r. **D** – ε difference between pol-β and T7-pol around the G∘T base pair in the active site. Data are means ± SD (n≥2 replica per model). *, $P < 0.05$; **, $P < 0.01$ (Student's t-test)

Free energy change of the wb-WC reaction from QM/MM US

The effects we observed might indicate a putative environmental contribution on the energetics of the wb-WC reaction, but we cannot estimate the extent of this contribution on the wb-WC equilibrium. Therefore, we calculated potential of mean force (PMF) of the wb-WC reaction in the studied molecular environments using QM/MM umbrella sampling (US) calculations. The PMF from the converged part of the US trajectories of the DNA model is shown on Fig. 4A and the converged PMFs from the rest of the systems are shown on Fig. S5. The only quantitative property we derived from the PMFs was the total free energy change of the wb-WC reaction $\Delta G_{wc} = \Delta G$(G*∘U WC) - ΔG(G∘U wb). ΔG_{wc} was used to evaluate the convergence of the US simulations by calculating PMF separately for consecutive batches of $\sim$ 20 ps.

Benchmark systems. In order to verify the validity of our approach, we first applied it to benchmark systems: G∘T base pair in the DNA heptamer duplex in water and a single G∘T base pair in benzene. ΔG_{wc} trajectories of the benchmark systems and the converged values of ΔG_{wc} are shown in Fig. 4B and Table 1, respectively. After initial fluctuations

from 4 to 10 kcal/mol, ΔG_{wc} of the DNA model converged to 7.2 ± 1.1 kcal/mol. Our calculations somewhat overestimate the experimental ΔG_{wc} range of 3.3-4.9 kcal/mol[16], but are consistent with the recent computational result of ∼ 6 kcal/mol derived from (ωB97X-D3/6-311G**)/AMBER calculations[18]. To assess the ability of our QM/MM setup to reproduce ΔG_{wc} dependence on ε observed in implicit solvent models, we applied it to a single G∘T base pair in explicit benzene environment ($\varepsilon = 2.3$). ΔG_{wc} in benzene converged to -0.5 ± 0.6 kcal/mol, which is consistent with the QM calculations in implicit solvent model of similar ε (Fig. S1). In both benchmark systems the positions of the minima on the PMF were not qualitatively altered compared to the reference reaction path (Fig. S5, Fig. 1C). From our benchmark calculations we conclude that our setup is able to faithfully estimate environmental effects on the wb-WC reaction, reaching qualitative agreement with experiments and higher levels of QM theory.

Table 1: ΔG_{wc} values from the US simulations

System	ΔG_{wc} [a]
DNA	7.2 ± 1.1
benzene	-0.5 ± 0.6
A-site closed	-0.7 ± 0.6
A-site abasic	11.2 ± 0.9
A-site open	6.6 ± 0.3
pol-β	-4.4 ± 1.6
T7-pol	7.2 ± 0.4

[a] Mean ΔG_{wc} ± standard deviation from the regions selected for convergence estimation on Fig. 4B-D, kcal/mol

Ribosomal A-site. We started the QM/MM US simulations of the A-site effects on the wb-WC reaction in the second codon-anticodon U∘G base pair in the models of closed and abasic A-site. ΔG_{wc} in the closed A-site converged to -0.7 ± 0.6 kcal/mol, while in the abasic A-site it converged to 11.2 ± 0.9 kcal/mol (Fig. 4C, Table 1). However, the abasic model is only a rough approximation of the open state of the A-site. To estimate effects of the rRNA nucleotides A1492 and A1493 more accurately, we created the "open" A-site model. In this model, harmonic restraints used in the previous studies[32] were applied to change

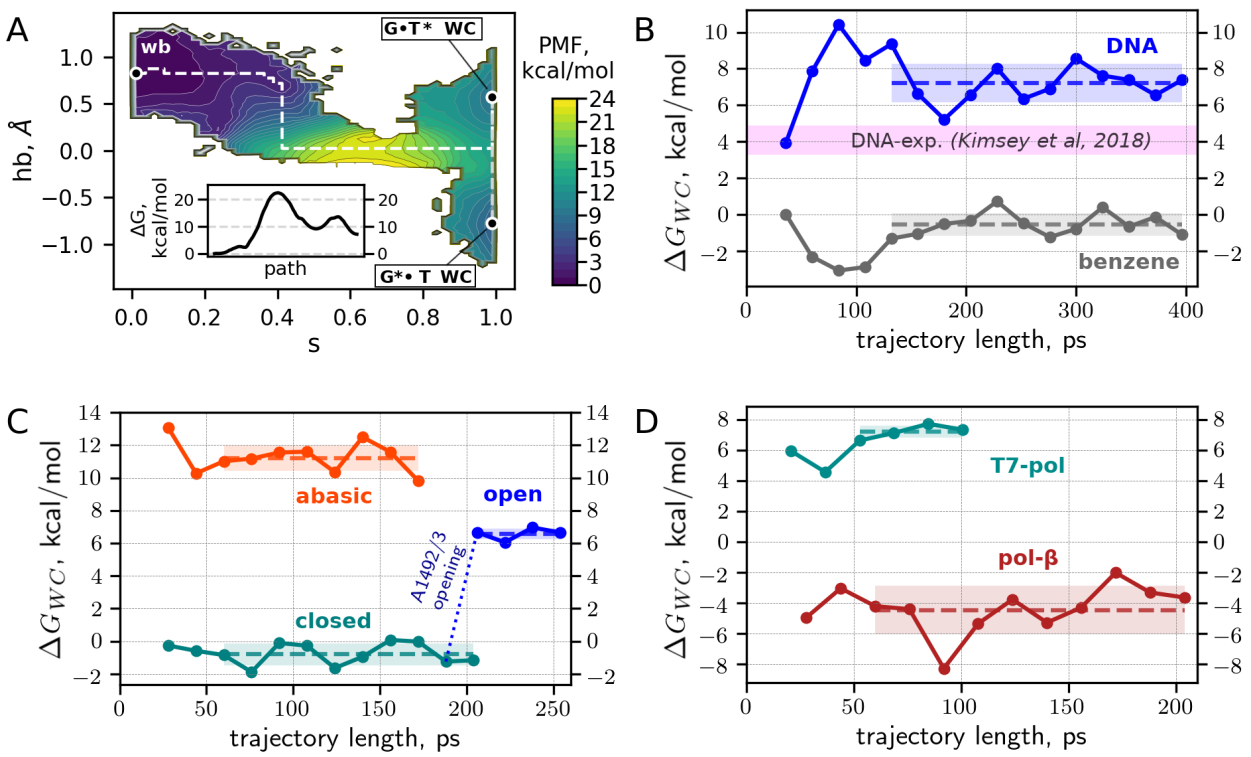


Figure 4: PMF from US calculations and its convergence. **A** – US-derived PMF of the wb-WC reaction in the DNA heptamer system. Black circles denote the minima corresponding to wb and WC structures. White dashed line denotes the minimal free energy path of the wb-WC reaction, from G∘T wb to G*∘T WC. The inset shows the free energy profile along the minimal free energy path. PMFs from the other six systems are shown on the Fig. S5. **B**–**D** – convergence of the ΔG_{wc} in benchmark systems (**B**), A-site models with G∘U in the middle codon position (**C**) and DNA polymerases (**D**). For each ΔG_{wc} trajectory, the shaded area and the dashed line of the corresponding color denote the standard deviation and mean of the set of batches selected for convergence evaluation. Magenta-shaded area in **B** denotes the range of ΔG_{wc} in DNA duplexes in solution measured by NMR[16].

A1492 and A1493 conformations from *out* to *in* (Fig. 2B). G530 remained in the closed-like conformation in the open model. These restraints were applied to the coordinates from the end part of the closed state US trajectories and maintained for approximately 70 ps in each US window. ΔG_{wc} in the open A-site converged to 6.6 ± 0.3 kcal/mol (Fig. 4C, Table 1). While in the abasic model the position of the wb minimum was not qualitatively altered compared to the reference reaction path ($s = 0$), in the closed and open models it shifted to higher values of the path variable ($s \approx 0.1$), suggesting a slight destabilization effect on the wobble geometry, likely exerted by G530 (Fig. S5). This effect was not enough to shift the

wb-WC equilibrium towards WC in the open model, indicating a critical role of A1492 and A1493 residues.

DNA polymerases. The previous QM/MM study have addressed effects of the DNA polymerase λ environment on the wb-WC reaction[18]. Here, we analyzed effects of two other DNA polymerases: low fidelity human DNA polymerase β (pol-β) and high-fidelity DNA polymerase from T7 virus (T7-pol). In both models, the wb-WC reaction was simulated in G∘T base pair in the closed state of a polymerase active site. ΔG_{wc} in pol-β converged to -4.4 ± 1.6 kcal/mol, indicating largely exoergic wb-WC reaction. This result corroborates previous structural studies, which reveal WC-like G∘T base pair in the closed active site of pol-β[17]. ΔG_{wc} in T7-pol converged to 7.2 ± 0.4 kcal/mol, resembling ΔG_{wc} value in the DNA duplex (Fig. 4B,D, Table 1). To the best of our knowledge, WC-like G∘T base pairs were never observed in the active site of T7-pol. Comparing PMFs from pol-β and T7-pol revealed a striking shift in the wb minimum position towards the higher value ($s \approx 0.2$) of the path variable in pol-β (Fig. S5). We attribute this shift to steric effects in pol-β that constrain a base pair geometry in the closed active site.

In sum, using QM/MM US calculations, we demonstrated the exoergic wb-WC reaction in G∘U at the middle codon-anticodon position in the closed A-site model, and the endoergic reaction in the open and abasic models. These results corroborate structural studies of the ribosome[20–24]. Decreased ε in the closed state of the decoding site can be one of the contributions to this effect. Shifts in the wb minimum position on the PMF in the closed and open models might indicate steric constraints on the base pair geometry. A similar effect was observed in pol-β active site, in which the wb-WC reaction was highly exoergic, in contrast to T7-pol, where the reaction equilibrium was not altered compared to the DNA duplex. More detailed studies are needed to delineate the environmental contributions on the properties of the wb-WC reaction.

The wb-WC reaction in the kinetic model of decoding in translation

The exoergic wb-WC reaction in the closed A-site requires reconsideration of the current model of decoding in translation, as the previously suggested "energy penalty" from tautomerization[20–23] cannot be the basis of selectivity. With the exoergic wb-WC reaction in equilibrium settings the error rate of G∘U mismatches in the second codon position would be close to 1, but the error rate is 10^{-2} to 10^{-3} instead[5–9]. To reconcile the exoergic wb-WC reaction with the model of decoding, the wb-WC reaction should be out of equilibrium in the closed A-site. To study the contribution of this reaction in decoding, we incorporated it into the model of initial selection.

Starting from the classical kinetic model of initial selection (Fig. S6), explained in details in[29], we introduced the wb-WC reaction by separating each near-cognate (nc) state into wb and WC states, leaving the cognate (c) branch unchanged (Fig. 5-A). Rate constants between nc-wb and nc-WC states correspond to the rate constants of the wb-WC reaction in a given environment of the A-site. To simplify the model for the sake of deriving analytical solutions, we assumed instant wb-shifted wb-WC equilibrium in states $C2$ (initial binding) and $C3$ (codon recognition in the open A-site). This approximation is well justified if, as suggested by our US calculations, the open state of the decoding site does not shift the equilibrium towards WC, and thus the equilibration time is negligible. In this approximation the wb-WC equilibrium in $C2$ does not affect the model and needed only for model completeness. The wb-WC reaction in $C4$ (codon recognition in the closed A-site) was modeled explicitly via its forward and reverse rate constants k_f, k_r. We used decoding rate constants measured previously at 20 °C[46,47]. The nc decoding rate constants in this set correspond not to a G∘U mismatch, but to the A∘C mismatch in the first position. This and other approximations preclude quantitative predictions of the error rate from the model. We analyzed effects of the wb-WC reaction on decoding that are not restricted to any particular combination of

the decoding rate constants.

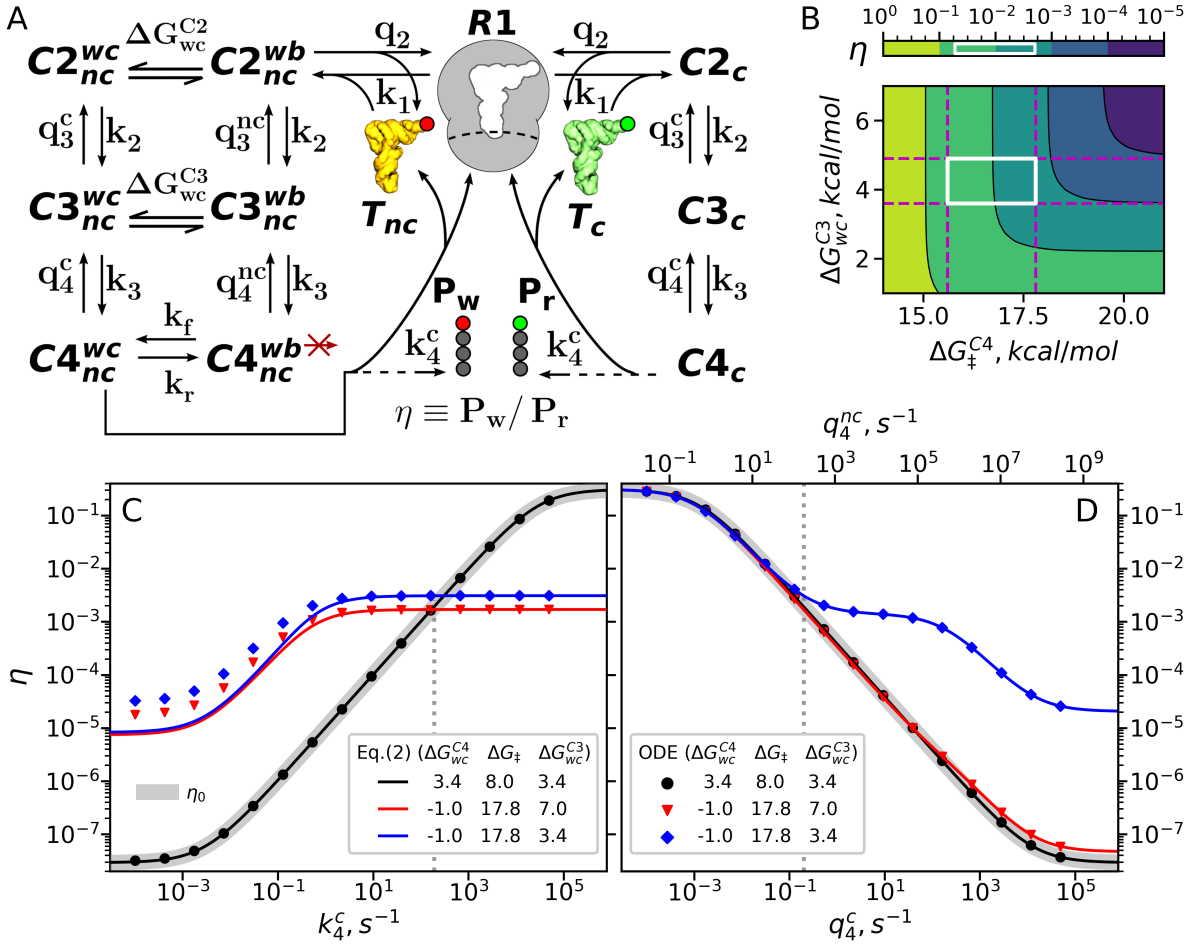


Figure 5: The wb-WC reaction in the kinetic model of decoding in translation. **A** – a scheme of the kinetic model of the initial selection in decoding with incorporated wb-WC reaction. Fig. S6 includes a detailed explanation of the decoding states and transitions. **B** – η as a function of $\Delta G_{\ddagger}$ and ΔG^{C3}_{wc}, calculated using Eq. (2) for $\Delta G^{C4}_{wc} = -1$ kcal/mol. The highlighted region corresponds to the experimental range of $\Delta G_{\ddagger}$ and ΔG_{wc} in RNA duplexes[16]. **C-D** error rate as a function of k^c_4 (**C**) and $q^{c/nc}_4$ (**D**), calculated by Eq. (A.1) for the classical system (η_0), and, for the specified wb-WC parameters, from numerical simulations (ODE) and by Eq. (2). $\eta(q^{c/nc}_4)$ and $\eta_0(q^{c/nc}_4)$ were obtained by varying q^c_4 and q^{nc}_4 simultaneously at constant q^c_4/q^{nc}_4 ratio. $\eta_0(k^c_4)$ was also obtained at constant k^c_4/k^{nc}_4 ratio. ΔG units in the plot legend are kcal/mol. The gray dotted vertical lines denote experimental values of k^c_4 and $q^{c/nc}_4$. The deviation of Eq. (2)-predicted η from η(ODE) for low values of k^c_4 in **C** is due to increased $C3$ state population when the wb-WC reaction approaches equilibrium.

Inspired by the previously applied numerical kinetic model of selection in replication by Kimsey et al.[16], we made the following assumptions: (i) nc-WC states are characterised with the same decoding rate constants as the cognate states, since the WC geometries are

assumed to be indistinguishable for the decoding site; (ii) GTPase activation rate constant (k_4) in $C4^{wb}_{nc}$ state is set to zero, assuming that the major contribution to this reaction rate comes from the nc-WC state, thus the contribution from nc-wb can be ignored; (iii) escape rate constants of the nc-wb states are taken from the classical nc states (q_i^{nc}), assuming that the cumulative escape rate is dominated by the nc-wb states, thus the nc→nc-wb rescaling can be neglected. Given the assumption (ii), the apparent near-cognate k_4:

$$k_4^{nc} = P_{wc} k_4^c \tag{1}$$

where P_{wc} is the population of the $C4^{wc}_{nc}$ state – a function of the equilibrium WC populations in $C3$ and $C4$ states, and constraints imposed by the decoding rates on the reaction kinetics in $C4$. Using the equation for product concentration in a first-order reversible reaction, we derived (see Appendix) the equation for the error rate induced by the wb-WC reaction:

$$\eta = \frac{[C4^{wc}_{nc}] + [C4^{wb}_{nc}]}{[C4_c]} \left(P^{eq}_{WC} + (P^{C3}_{WC} \frac{q_4^{nc}}{k_4^c + q_4^c} - P^{eq}_{WC}) \exp\left(-\frac{k_f + k_r}{k_4^c + q_4^c}\right) \right) \tag{2}$$

where P^{eq}_{WC} is the equilibrium WC population in $C4$ for a given (k_f, k_r), and P^{C3}_{WC} is the equilibrium WC population in $C3$. For convenience, the variables P^{eq}_{WC}, P^{C3}_{WC} and k_f are expressed below in terms of free energy differences ΔG^{C4}_{wc}, ΔG^{C3}_{wc} and $\Delta G_{\ddagger}$ (activation free energy), respectively, using Eq. (S.5) and Eq. (S.6) at standard conditions.

First, we analyzed how ΔG^{C3}_{wc} and $\Delta G_{\ddagger}$ of the exoergic ($\Delta G^{C4}_{wc} = -1$ kcal/mol) wb-WC reaction affect η predicted from Eq. (2) (Fig. 5B). For the range of ΔG_{wc} and $\Delta G_{\ddagger}$ in RNA duplexes in solution reported in[16], η overlaps with the *in vitro* error range of G∘U mismatches ($\sim 10^{-2} - 10^{-4}$)[5–9]. However, this result clearly cannot be interpreted as a validation of Eq. (2). In order to test Eq. (2), the decoding rate constants for G∘U mismatches should be obtained, and $\Delta G_{\ddagger}$ in the decoding site accurately estimated.

Next, we addressed the dependence of η on the decoding rate constants k_4^c (GTPase

activation) and $q_4^{c/nc}$ (cognate and near-cognate escape rate constants of the $C4$ state). These rate constants determine the deviation of the open↔closed transition in decoding from equilibrium conditions, thereby limiting the accuracy of decoding in the classical model[27,28]. Eq. (2) suggests two possible kinetic regimes of the wb-WC reaction for the case of positive ΔG_{wc}^{C3}. In the "fast" regime ($(k_f + k_r) >> (k_4^c + q_4^c)$), kinetics of the wb-WC reaction and P_{WC}^{C3} are irrelevant, and $P_{wc} = P_{WC}^{eq}$. For $P_{WC}^{eq} = k_4^{nc}/k_4^c$ (corresponds to $\Delta G_{wc}^{C4} \approx 3.4$ kcal/mol), the fast regime is equivalent to the classical error η_0 (Eq. (A.1)) in both k_4^c and $q_4^{c/nc}$ dependencies, which is confirmed by numerical calculations (Fig. 5C-D). Therefore, our model satisfies the correspondence principle[50]: it reduces to the classical induced-fit model when the wb-WC reaction is at equilibrium in the pre-chemistry step of decoding.

A more intriguing and relevant given the predicted wb-WC parameters is the "slow" regime, where the wb-WC equilibrium in $C4$ is shifted to the WC geometry ($k_f > k_r$), but the kinetics of the reaction is restricted by the decoding rates. To visualize the slow regime, we chose wb-WC parameters ($\Delta G_{wc}^{C4} = -1$ kcal/mol, $\Delta G_{\ddagger} = 17.8$ kcal/mol), which approximately correspond to the predicted ΔG_{wc}^{C4} (Table 1) and experimentally observed $\Delta G_{\ddagger}$ in RNA duplexes[16]. For $k_f < k_4^c$, Eq. (2) predicts a virtually flat $\eta(k_4^c)$ curve, as confirmed by numerical solutions, and ΔG_{wc}^{C3} affects the offset of the curves (Fig. 5C). The reason for the flat $\eta(k_4^c)$ curve in the slow regime is derived in Appendix and visualized on Fig. S7: the linear approximations of P_{wc} and $\frac{[C4_{nc}]}{[C4_c]}$ have inverse dependency on k_4^c, which cancels in η. Equilibration of the exoergic wb-WC reaction in $C4$ with unfavorable contribution to the accuracy counteracts the equilibration of the open→closed transition with favorable contribution, thereby constraining the decoding accuracy.

In $\eta(q_4^{c/nc})$, the slow kinetic regime with a very low P_{WC}^{C3} ($\Delta G_{wc}^{C3} = 7.0$ kcal/mol) closely follows η_0. However, the same kinetic regime with $\Delta G_{wc}^{C3} = 3.4$ kcal/mol results in an almost flat curve for $q_4^{nc} > k_4^c > q_4^c$ (Fig. 5D). This effect is explained by the kinetic partitioning term in $P_{WC}^{C3} \frac{q_4^{nc}}{k_4^c + q_4^c}$, which grows proportionally with q_4^{nc} and cancels the classical equilibration process. In sum, Eq. (2), confirmed by the numerical calculations, predicts that high $\Delta G_{\ddagger}$

in $C4$ causes a flat $\eta(k_4^c)$ curve for $k_f < k_4^c$, while high P_{WC}^{C3} causes an almost flat $\eta(q_4^{c/nc})$ for $q_4^{nc} > k_4^c > q_4^c$. Two potentially independent parameters of the wb-WC reaction constrain the decoding accuracy of the G∘U mismatch recognition in translation, which might have significant implications for the evolutionary optimization of decoding.

To explore this further, we considered a cumulative error rate of translation, η_c, as a sum of η contributions from mismatches with slow, WC-shifted wb-WC transitions (e.g., G∘U mismatches) and with "classical" η_0-like behavior. In terms of our model (Fig. 5A), the latter can be interpreted as mismatches having fast endoergic transition to the WC geometry in $C4$, or mismatches whose GTPase activation rate is governed by low WC-independent rate constant (i.e. $k_4^{wb} \neq 0, \frac{k_4^{wb}}{k_4^c} = const$), that is ignored when considering G∘U mismatches. Recent studies show that the error hotspots (primarily G∘U mismatches) have error rate of $\sim 10^{-2} - 10^{-4}$ *in vitro*, and the error rate of other mismatches span a range of $10^{-4} - 10^{-7}$[5–9]. Based on these findings, we approximated η_c as sum of η of the slow kinetic regime ($\Delta G_{wc}^{C4} = -1$ kcal/mol, $\Delta G_{\ddagger} = 17.8$ kcal/mol, $\Delta G_{wc}^{C3} = 3.4$ kcal/mol) and a range of η_0 curves with rescaled k_4^{nc}/k_4^c ratio that results in η_0 range of $10^{-3} - 10^{-8}$ at the experimental values of the decoding rate constants. We calculated η_c as a function of k_4^c and $q_4^{c/nc}$. The resulting η_c curves are shown as dashed lines on Fig. S8. η_c is governed by η of the G∘U mismatch for all values of $q_4^{c/nc}$, and for decreasing values of k_4^c.

The evolutionary optimization of the values of k_4^c and $q_4^{c/nc}$ cannot be explained solely by η or η_c, which are monotonic functions of these rate constants. To explore possible reasons for the evolutionary optimization of these rate constants to the values observed in the experiments, we looked for extrema in the gradient of η_c, $\nabla\eta_c$. $\nabla\eta_c$ determines the noise (i.e. standard deviation) of error rate η_c, given that η_c is monotonic. However, calculating the noise of $\eta_c(k_4^c)$ and $\eta_c(q_4^{c/nc})$ would require knowing the variance of k_4^c and $q_4^{c/nc}$. Therefore, for this illustrative purpose, we restricted to the analysis of $\nabla\eta_c$. Numerical gradients $\nabla\eta_c$ were calculated for each η_c curve and shown on Fig. S8 as solid lines. At η_0 contribution of 10^{-5}, the experimental value of k_4^c is at the local minimum of $\nabla\eta_c$ (Fig. S8). We selected this

η_0 parameter to visualize the cognate rate of decoding R_{cog}, η_c and $\nabla\eta_c$ as 2D functions of k_4^c and $q_4^{c/nc}$ (Fig. 6). Both rate constants for a wide range do not significantly contribute to R_{cog} (Fig. 6A, Fig. S9). The exoergic wb-WC reaction constrains η_c at a plateau (Fig. 6B): for the ribosome to significantly improve the accuracy, the decoding rate needs to be reduced. In contrast, for a classical mismatch, the ribosome could reduce the error rate at least 1000-fold without significantly affecting the decoding rate (Fig. S9A).

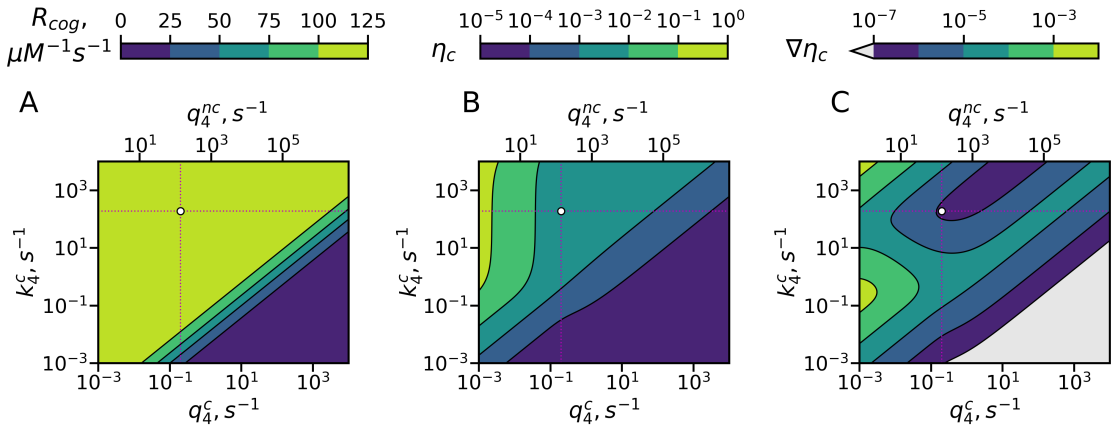


Figure 6: Optimization landscape of decoding constrained by η of G∘U mismatches. The cognate rate of decoding (R_{cog}, **A**), cumulative error rate (η_c, **B**) and the gradient of cumulative error rate ($\nabla\eta_c$, **C**) were calculated as functions of k_4^c and $q_4^{c/nc}$. Each function was calculated as a sum of corresponding contributions from the slow regime of the wb-WC reaction ($\Delta G_{wc}^{C4} = -1$ kcal/mol, $\Delta G_{\ddagger} = 17.8$ kcal/mol, $\Delta G_{wc}^{C3} = 3.4$ kcal/mol) and a "classical" mismatch with rescaled k_4^{nc}/k_4^c ratio that results in $\eta_0 = 10^{-5}$ at the experimental values of the decoding rate constants. Dotted lines and white circle at the intersection denote the experimental values of k_4^c and $q_4^{c/nc}$.

In the cumulative approximation, the ribosome (the experimental values of decoding rate constants) is positioned in a region of local minimum of $\nabla\eta_c$ (Fig. 6C). It illustrates the possibility of the evolutionary optimization for the minimal gradient, and thus for the minimal noise of error rate. For the case of η of a G∘U mismatch only, this position is shifted from the minimum (Fig. S9B). Consideration of the cumulative error of decoding would likely be important to understand the evolutionary optimization of the ribosome, but requires more experimental information and more realistic approximations.

It is also worth noting that our model does not contradict the experiments with varying Mg^{2+} concentrations. In their experiments Zhang et al.[51] reported that Mg^{2+} concentration

affected q_2, resulting in linear rate-accuracy trade-offs. For all considered wb-WC parameters, our model also predicts linear rate-accuracy trade-offs by varying q_2 (Fig. S10).

CONCLUDING REMARKS

Structural studies revealed stabilization of the WC geometry of the G∘U base pairs in the closed ribosomal A-site[20–24], but the mechanism of this phenomenon has been lacking a physichochemical explanation and its implications for the decoding mechanism remained unexplored. Here we addressed these problems using computational and theoretical approaches.

QM/MM umbrella sampling simulations show that closing of the rRNA residues A1492 and A1493 in the decoding site shifts the equilibrium in the wb-WC reaction to the WC geometry at the middle codon-anticodon position. This effect may partly be explained by the decreased polarity of the closed decoding site. The model of the ribosomal A-site used in our simulations is of course a simplification of the real conditions that may affect the wb-WC properties during the decoding. The major simplifications are: the model is only a part of the ribosome structure, with an outer shell that was restrained during all simulations; absent Mg^{2+} ions in the model; the "open" model differed from the "closed" model only by conformations of A1492 and A1493 residues. However, we argue that this model is able to capture the major effects of the open→closed transition on the wb-WC equilibrium. Distant structural elements and their change during the 30S domain closure are unlikely to dramatically affect the wb-WC reaction, and Mg^{2+} ions have not been reported to directly interact with the second codon-anticodon position. The role of G530, which was not affected in the open model, is however not fully explored in our study. It is remained to be investigated how the wb-WC reaction is affected in the first and third codon-anticodon positions which were not addressed in this study.

We also investigated the effects of the closed active site environment of DNA polymerases

β and T7 on the wb-WC reaction in G∘T. The wb-WC reaction on pol-β was exoergic, corroborating previous structural studies[17]. The wb-WC reaction equilibrium in T7-pol was not affected compared to the DNA duplex in water.

We introduced the wb-WC reaction into the classical model of initial selection in translation and derived analytical solutions for the decoding accuracy governed by the wb-WC reaction. The analytical solutions fit the numerical calculations and explain the decoding accuracy of G∘U mismatch recognition in terms of the wb-WC reaction parameters in open and closed states of the A-site. This model is consistent with the WC geometry of G∘U observed in the structural studies at equilibrium conditions ($k_4^c \to 0$)[20–24]. Since a set of decoding rate constants for a G∘U mismatch and the accurate ΔG_{wc}^{C3} and $\Delta G_{\ddagger}$ are not currently available, our model here does not serve a purpose of accurate error rate prediction. It reveals possible constraints on the codon-anticodon decoding imposed by the exoergic wb-WC tautomerization reaction: the equilibration of the reaction counteracts the equilibration of the open↔closed transition. These constraints provide a solution to the long-standing problem of seemingly suboptimal decoding mechanism pictured by the induced-fit model[26]. We speculate that the model derived here for the codon-anticodon decoding in translation can also be applicable for the base pair recognition in DNA replication. We discuss specific predictions of our model, as well as relations to the previous models and available experimental data in Supplementary Discussion. We hope that our theoretical and computational results might be a next step towards a more complete model of substrate recognition.

Acknowledgement

We thank Prof. Andrew Torda for helpful comments on the manuscript. A.K. thanks Prof. Dmytro Hovorun for an inspiration in the problem of tautomerism of nucleobases. The work was supported by a fellowship from Deutscher Akademischer Austauschdienst (DAAD) to A.K. We are grateful for computational support by Norddeutscher Verbund für Hoch- und

Höchstleistungsrechnen (HLRN) (allocation code hhc00026), and by Hummel cluster at the University of Hamburg.

Supporting Information Available

Detailed methods, a detailed discussion of the new model, derivations of the analytical solutions of the model, Tables S1-S3 and Figures S1-S12 are available in the Supplementary Information.

Data availability

Optimized base pair geometries, coordinate frames of the NEB trajectory, initial system topologies and coordinates in the QM/MM US simulations, as well an example setup of the QM/MM US simulations are available on Figshare:

https://doi.org/10.6084/m9.figshare.c.5178074.v1

Code availability

Python and Tcl scripts for the umbrella sampling setup, dielectric permittivity calculations, kinetic modeling and visualizations are available on GitHub:

https://github.com/and-kaz/wbwc_paper

References

(1) Santos, M.; Pereira, P. M.; Varanda, A. S.; Carvalho, J.; Azevedo, M.; Mateus, D. D.; Mendes, N.; Oliveira, P.; Trindade, F.; Pinto, M. T.; Bordeira-Carriço, R.; Carneiro, F.; Vitorino, R.; Oliveira, C.; Santos, M. A. Codon misreading tRNAs promote tumor growth in mice. *RNA Biology* **2018**, *15*, 773–786.

(2) Ke, Z.; Mallik, P.; Johnson, A. B.; Luna, F.; Nevo, E.; Zhang, Z. D.; Gladyshev, V. N.; Seluanov, A.; Gorbunova, V. Translation fidelity coevolves with longevity. *Aging Cell* **2017**, *16*, 988–993.

(3) Hershberg, R. Mutation—the engine of evolution: Studying mutation and its role in the evolution of bacteria. *Cold Spring Harbor Perspectives in Biology* **2015**, *7*, a018077.

(4) Nei, M.; Suzuki, Y.; Nozawa, M. The Neutral Theory of Molecular Evolution in the Genomic Era. *Annual Review of Genomics and Human Genetics* **2010**, *11*, 265–289.

(5) Pernod, K.; Schaeffer, L.; Chicher, J.; Hok, E.; Rick, C.; Geslain, R.; Eriani, G.; Westhof, E.; Ryckelynck, M.; Martin, F. The nature of the purine at position 34 in tRNAs of 4-codon boxes is correlated with nucleotides at positions 32 and 38 to maintain decoding fidelity. *Nucleic acids research* **2020**, *48*, 6170–6183.

(6) Garofalo, R.; Wohlgemuth, I.; Pearson, M.; Lenz, C.; Urlaub, H.; Rodnina, M. V. Broad range of missense error frequencies in cellular proteins. *Nucleic acids research* **2019**, *47*, 2932–2945.

(7) Mordret, E.; Dahan, O.; Asraf, O.; Rak, R.; Yehonadav, A.; Barnabas, G. D.; Cox, J.; Geiger, T.; Lindner, A. B.; Pilpel, Y. Systematic Detection of Amino Acid Substitutions in Proteomes Reveals Mechanistic Basis of Ribosome Errors and Selection for Translation Fidelity. *Molecular Cell* **2019**,

(8) Manickam, N.; Nag, N.; Abbasi, A.; Patel, K.; Farabaugh, P. J. Studies of translational misreading in vivo show that the ribosome very efficiently discriminates against most potential errors. *RNA* **2014**, *20*, 9–15.

(9) Zhang, Z.; Shah, B.; Bondarenko, P. V. G/U and Certain Wobble Position Mismatches as Possible Main Causes of Amino Acid Misincorporations. *Biochemistry* **2013**, *52*, 8165–8176.

(10) Joshi, K.; Cao, L.; Farabaugh, P. J. The problem of genetic code misreading during protein synthesis. *Yeast* **2019**, *36*, 35–42.

(11) Watson, J. D.; Crick, F. H. Genetical implications of the structure of deoxyribonucleic acid. *Nature* **1953**, *171*, 964–967.

(12) Watson, J. D.; Crick, F. H. The structure of DNA. *Cold Spring Harbor symposia on quantitative biology* **1953**, *18*, 123–131.

(13) Kimsey, I. J.; Petzold, K.; Sathyamoorthy, B.; Stein, Z. W.; Al-Hashimi, H. M. Visualizing transient Watson-Crick-like mispairs in DNA and RNA duplexes. *Nature* **2015**, *519*, 315–320.

(14) Brovarets, O. O.; Hovorun, D. M. Physicochemical mechanism of the wobble DNA base pairs Gu·Thy and Ade·Cyt transition into the mismatched base pairs Gua*·Thy and Ade·Cyt* formed by the mutagenic tautomers. *Ukrainica Bioorganica Acta 2* **2009**, *7*, 12–18.

(15) Brovarets, O. O.; Hovorun, D. M. How many tautomerization pathways connect Watson-Crick-like G*·T DNA base mispair and wobble mismatches? *J Biomol Struct Dyn* **2015**, 1–19.

(16) Kimsey, I. J.; Szymanski, E. S.; Zahurancik, W. J.; Shakya, A.; Xue, Y.; Chu, C. C.; Sathyamoorthy, B.; Suo, Z.; Al-Hashimi, H. M. Dynamic basis for dG· dT misincorporation via tautomerization and ionization. *Nature* **2018**, *554*, 195–201.

(17) Koag, M. C.; Nam, K.; Lee, S. The spontaneous replication error and the mismatch discrimination mechanisms of human DNA polymerase β. *Nucleic Acids Research* **2014**, *42*, 11233–11245.

(18) Li, P.; Rangadurai, A.; Al-Hashimi, H. M.; Hammes-Schiffer, S. Environmental Effects on Guanine-Thymine Mispair Tautomerization Explored with Quantum Mechanical/Molecular Mechanical Free Energy Simulations. *Journal of the American Chemical Society* **2020**, jacs.0c03774.

(19) Maximoff, S. N.; Kamerlin, S. C. L.; Florian, J. DNA Polymerase λ Active Site Favors a Mutagenic Mispair Between the Enol Form of Deoxyguanosine Triphosphate Substrate and the Keto Form of Thymidine Template. A Free Energy Perturbation Study. *The Journal of Physical Chemistry B* **2017**,

(20) Demeshkina, N.; Jenner, L.; Westhof, E.; Yusupov, M.; Yusupova, G. A new understanding of the decoding principle on the ribosome. *Nature* **2012**, *484*, 256–259.

(21) Demeshkina, N.; Jenner, L.; Westhof, E.; Yusupov, M.; Yusupova, G. New structural insights into the decoding mechanism: translation infidelity via a G.U pair with Watson-Crick geometry. *FEBS Lett* **2013**, *587*, 1848–1857.

(22) Rozov, A.; Demeshkina, N.; Westhof, E.; Yusupov, M.; Yusupova, G. Structural insights into the translational infidelity mechanism. *Nat Commun* **2015**, *6*, 7251.

(23) Rozov, A.; Wolff, P.; Grosjean, H.; Yusupov, M.; Yusupova, G.; Westhof, E. Tautomeric G•U pairs within the molecular ribosomal grip and fidelity of decoding in bacteria. *Nucleic Acids Research* **2018**,

(24) Loveland, A. B.; Demo, G.; Grigorieff, N.; Korostelev, A. A. Ensemble cryo-EM elucidates the mechanism of translation fidelity. *Nature* **2017**, *546*, 113–117.

(25) Ogle, J. M.; Murphy IV, F. V.; Tarry, M. J.; Ramakrishnan, V. Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell* **2002**, *111*, 721–732.

(26) Johansson, M.; Lovmar, M.; Ehrenberg, M. Rate and accuracy of bacterial protein synthesis revisited. *Current Opinion in Microbiology* **2008**, *11*, 141–147.

(27) Wohlgemuth, I.; Pohl, C.; Mittelstaet, J.; Konevega, A. L.; Rodnina, M. V. Evolutionary optimization of speed and accuracy of decoding on the ribosome. *Philosophical Transactions of the Royal Society B: Biological Sciences* **2011**, *366*, 2979–2986.

(28) Savir, Y.; Tlusty, T. The Ribosome as an Optimal Decoder: A Lesson in Molecular Recognition. *Cell* **2013**, *153*, 471–479.

(29) Pavlov, M. Y.; Ehrenberg, M. Substrate-Induced Formation of Ribosomal Decoding Center for Accurate and Rapid Genetic Code Translation. *Annual Review of Biophysics* **2018**, *47*, 525–548.

(30) Mallory, J. D.; Kolomeisky, A. B.; Igoshin, O. A. Trade-Offs between Error, Speed, Noise, and Energy Dissipation in Biological Processes with Proofreading. *Journal of Physical Chemistry B* **2019**, *123*, 4718–4725.

(31) Satpati, P.; Aqvist, J. Why base tautomerization does not cause errors in mRNA decoding on the ribosome. *Nucleic Acids Res* **2014**, *42*, 12876–12884.

(32) Zeng, X.; Chugh, J.; Casiano-Negroni, A.; Al-Hashimi, H. M.; Brooks Iii, C. L. Flipping of the Ribosomal A-Site Adenines Provides a Basis for tRNA Selection. *Journal of Molecular Biology* **2014**, *426*, 3201–3213.

(33) Stewart, J. J. Optimization of parameters for semiempirical methods VI: More modifications to the NDDO approximations and re-optimization of parameters. *Journal of Molecular Modeling* **2013**, *19*, 1–32.

(34) Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018**, *8*.

(35) Stewart, J. J. Mopac2016. 2016.

(36) Freudenthal, B. D.; Beard, W. A.; Shock, D. D.; Wilson, S. H. Observing a DNA polymerase choose right from wrong. *Cell* **2013**, *154*, 157.

(37) Doublié, S.; Tabor, S.; Long, A. M.; Richardson, C. C.; Ellenberger, T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **1998**, *391*, 251–258.

(38) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

(39) Huang, J.; Mackerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry* **2013**, *34*, 2135–2145.

(40) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain $\chi1$ and $\chi2$ Dihedral Angles. *Journal of Chemical Theory and Computation* **2012**, *8*, 3257–3273.

(41) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D. Impact of 2′-hydroxyl sampling on the conformational properties of RNA: Update of the CHARMM all-atom additive force field for RNA. *Journal of Computational Chemistry* **2011**, *32*, 1929–1943.

(42) Phillips, J. C. et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *Journal of Chemical Physics* **2020**, *153*, 044130.

(43) Neumann, M. The dielectric constant of water. Computer simulations with the MCY potential. *The Journal of Chemical Physics* **1984**, *82*, 5663–5672.

(44) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in free energy space. *Journal of Chemical Physics* **2007**, *126*, 054103.

(45) Pape, T.; Wintermeyer, W.; Rodnina, M. V. Conformational switch in the decoding region of 16S rRNA during aminoacyl-tRNA selection on the ribosome. *Nature Structural Biology* **2000**, *7*, 104–107.

(46) Rudorf, S.; Thommen, M.; Rodnina, M. V.; Lipowsky, R. Deducing the kinetics of protein synthesis in vivo from the transition rates measured in vitro. *PLoS Comput Biol* **2014**, *10*, e1003909.

(47) Gromadski, K. B.; Daviter, T.; Rodnina, M. V. A uniform response to mismatches in codon-anticodon complexes ensures ribosomal fidelity. *Molecular Cell* **2006**, *21*, 369–377.

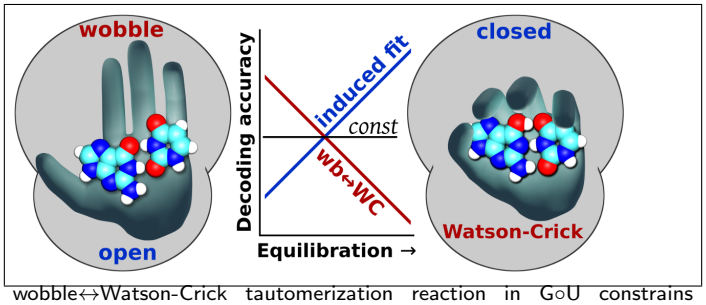
(48) Satpati, P.; Sund, J.; Åqvist, J. Structure-based energetics of mRNA decoding on the ribosome. *Biochemistry* **2014**, *53*, 1714–1722.

(49) Nomura, K.; Hoshino, R.; Shimizu, E.; Hoshiba, Y.; Danilov, V. I.; Kurita, N. DFT Calculations on the Effect of Solvation on the Tautomeric Reactions for Wobble Gua-Thy and Canonical Gua-Cyt Base-Pairs. *Journal of Modern Physics* **2013**, *4*, 422–431.

(50) Bohr, N. Über die Serienspektra der Elemente. *Zeitschrift für Physik* **1920**, *2*, 423–469.

(51) Zhang, J.; Pavlov, M. Y.; Ehrenberg, M. Accuracy of genetic code translation and its orthogonal corruption by aminoglycosides and Mg^{2+} ions. *Nucleic Acids Research* **2018**, *46*, 1362–1374.

Graphical TOC Entry



wobble↔Watson-Crick tautomerization reaction in G∘U constrains the accuracy of decoding at the ribosome by counteracting the favorable contribution from the induced fit.